心理科学进展 2022, Vol. 30, No. 12, 2619–2627
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2022.02619

·研究构想(Conceptual Framework)·

# 智能组织中的人机协同决策：基于人机内部兼容性的研究探索*

何贵兵 陈 诚 何泽桐 崔力丹 陆嘉琦 宣泓舟 林 琳

(浙江大学心理与行为科学系, 杭州 310028)

**摘 要** 智能时代已然来临。智能技术的发展正加速推动组织的智能化进程。越来越多的企业在生产和管理中采用智能技术以提高竞争力。在此背景下，人机协同工作日益普遍，人机协同决策成为新型组织决策方式。然而，当前智能组织中的人机协同决策还面临信任度低、可控性低、透明度低、协同度低等一系列问题，它们阻碍了决策质量、效率和体验的提升。本项目认为，人机兼容性，特别是人机内部兼容性，如认知兼容性、情感兼容性、价值兼容性等，或是影响人机协同决策绩效的根本原因。因此，本项目基于人机内部兼容性理论视角，综合采用决策心理学、认知科学、组织行为学等多学科理论与方法，通过一系列现场研究和模拟实验，力图揭示人机协同决策中存在的问题及其成因，探究人机协同决策中内部兼容性的影响因素和作用机制，进而提出若干人机协同决策优化方法。项目研究成果将有助于促进人机协同决策理论与人机兼容性理论的发展，提升人机协同决策绩效，推进组织决策的智能化进程。

**关键词** 智能组织，人机协同决策，内部兼容性，认知兼容性，情感兼容性，价值兼容性
**分类号** B849

## 1 问题提出

智能时代已然来临。伴随着新一轮科技革命和产业革命，人类社会正从工业化时代、信息化时代迈入智能化时代。大数据与智能革命正深刻改变着人们的生产、生活和学习方式，重新定义人们的未来。在智能技术的推动下，智慧城市、智慧政务、智慧农业、智慧医疗、智能教育、智慧司法、智能交通等各领域的智能化变革日新月异。与此同时，越来越多的企业认识到智能技术对提高组织竞争力的重要性，因而通过制定战略、加大投入、培养人才等途径将智能技术应用于生产、经营、管理等各环节，企业组织的智能化趋势越来越明显。毕马威和阿里研究院(2019)在研究报告《百年跃变：浮现中的智能化组织》中将引入智能技术以提升运行效能的组织称为智能组织(intelligent organization)，并视其为未来组织的发展方向。2018 年，麦肯锡的一项关于企业人工智能采用情况的调查发现，将近 50%的受访者表示其公司已经在业务流程中至少使用了一项人工智能技术，另有 30%的企业正在试用人工智能技术。

组织智能化引发了工作方式、管理模式乃至人机关系的深刻变革。首先，智能体(intelligent agent，包括智能软件和硬件)的应用极大提升了业务流程的自动化水平，工作方式由人主导转向人机协作甚至由智能体主导。其次，智能体的应用使组织管理模式发生深刻变革。数据、算法和算力成为重要的管理资源，扁平化、自适应、协同、精准、高效成为重要的管理特征。另外，由于新一代智能体具有更强的自主学习能力与问题处理能力，其与人类工作者的关系亦被重新定义(许为，葛列众，2020)。在组织中，智能体不再只是人类的“工具”，其地位与人类成员日渐平等，真正成为人类的“伙伴”或“同事”，甚至扮演“领

收稿日期：2021-12-31
* 国家自然科学基金项目(72071178)。
通信作者：何贵兵, E-mail: gbhe@zju.edu.cn

导”、“管理者”的角色(Christoffersen & Woods, 2002; Madhavan & Wiegmann, 2007; Wynne & Lyons, 2018)。越来越多的人类员工将与智能体组成人机团队(human-agent team)。此类团队由人机构成，通过人机间互动协作共同执行组织任务，从而发挥各自优势，更好实现组织任务目标(Scheutz et al., 2017)。

智能体在组织管理决策中也扮演着越来越重要的角色。以往，组织决策质量主要取决于人类决策者的认知加工能力。然而，在大数据时代，人们需要面对海量信息，其认知加工能力的局限性更显突出，这使得智能体越来越多地被赋予决策者职能，以减轻人的认知负荷，并促进决策优化(Chamorro-premuzic & Ahmetoglu, 2016; Davenport & Ronanki, 2018)。智能体在组织决策中承担着辅助与主导两种角色。作为决策辅助者，智能体通过机器学习分析海量信息、挖掘潜在规律，为组织做出更优决策提供数据支持或决策建议(刘文川 等, 2019)。作为决策主导者，智能体被赋予更高决策权限，能够主导组织的分配决策(Ötting & Maier, 2018)、投资决策(Kaplan & Haenlein, 2019)，甚至是人员招聘、培训等人事决策(Dineen et al., 2004; Langer et al., 2016; Lee, 2018; Naim et al., 2016)。决策智能体不仅能为企业解决决策实时性要求高、数据量巨大的问题(Harms & Han, 2019)，同时也能促使决策程序标准化，提升决策公平感，减少组织冲突，减轻领导者的管理压力(Ötting & Maier, 2018)。

然而，在大多数管理决策中，人类决策者的智慧和控制需求仍不可或缺，因此人机协同决策(Human-agent collaborative decision-making)，即由人类决策者与智能体共同完成决策，成为智能组织中更常见的决策方式。当然，在不同决策任务中，或在决策过程不同阶段中，人机双方的参与程度和协作水平会存在差异。在相同决策任务下，不同水平的人机协同可能产生不同质量的决策。

一系列研究表明，当前组织中人机协同决策的效果并不尽如人意，人机协同决策还存在不少问题。一是人类对智能体决策者的信任度与接受度低，严重阻碍了智能体在组织决策中的广泛应用(Ghislieri et al., 2018)；二是智能体的决策通常基于高度复杂和不透明的算法，人类往往难以知晓其决策规则及其内在含义，以至于人无法在智能体出现异常时及时采取应对措施(Dewhurst & Willmott, 2014; Lee, 2018; Ötting & Maier, 2018)；三是人机协同决策时，人类与智能体的决策权分配也是一大难题(Parry et al., 2016)；四是智能体决策所遵循的价值规则可能不符合人类决策者价值观和道德准则，甚至发生冲突(Yokoi & Nakayachi, 2021)。

显然，人机协同决策的质量、效率与主观体验均有待提升(Burton et al., 2020)。我们认为，出现上述问题的原因，或与人机兼容性(human-agent compatibility)有关。Bubb (1993)曾将人机兼容性分为外部兼容性(outer compatibility)与内部兼容性(inner compatibility)。外部兼容性反映人机双方在外部交互界面上的匹配度，内部兼容性关注双方在内部系统上的匹配度。我们提出，人类决策者与智能体间的内部兼容性，尤其是认知兼容性、情感兼容性和价值兼容性，是影响人机协同决策绩效的深层次原因。因此，有必要从人机内部兼容性视角出发，探讨人机协同决策中存在的问题及成因，分析认知兼容性、情感兼容性、价值兼容性对人机协同决策过程和决策绩效的潜在影响，并针对性地提出若干优化人机协同决策的方法，以提高组织的人机协同决策质量、效率和体验，更好推进组织决策智能化进程。

## 2 研究现状

本项目拟考察智能组织中人机内部兼容性与人机协同决策过程及决策绩效的关系。虽然以往有少量研究涉及认知兼容、人机信任与人机互动的关系，但探讨人机协同决策中内部兼容性及其各成分的作用，仍是个新课题。考虑到“人人团队决策研究”与“人机协同决策研究”可能存在类比关系，“人机协同关系”与“人机协同决策”亦有密切关联，因此，我们主要介绍“人人团队决策”和“人机协同关系”领域的相关研究和理论成果。其中，前者包括团队共享信息理论、团队共享心智模型与团队决策认知过程的 STC 机制等；后者包括人机兼容性理论、人机信任理论、技术接受模型等。上述内容对本项目理论构思的形成具有启发和借鉴意义。

### 2.1 关于人人团队决策的研究

(1)团队共享信息理论

组织的重要决策通常由团队成员共同完成

(Larson et al., 1996)。Stasser 和 Titus (1985)将团队成员决策前所持有的相关信息分为两类：共享信息(Shared information)和非共享信息(Unshared information)。前者指全体成员共有的信息；后者指为各成员独有的信息。决策开始前，团队成员所掌握的决策信息各有不同，其初始偏好亦存在差异。高质量的团队决策要求各方成员在决策过程中充分分享信息，尤其是非共享信息(Lam & Schaubroeck, 2011)。然而，Stasser 和 Titus (1987)发现，当进行团队决策时，团队成员往往更关注共享信息，较少讨论非共享信息，并将这一现象称为共享信息偏差 (Shared information bias)。共享信息偏差主要表现为三个方面：(1)相较于非共享信息，共享信息在讨论过程中被团队成员更为频繁地提及；(2)相较于非共享信息，共享信息被更早地引入团队讨论；(3)讨论结束后，团队成员对共享信息的回忆率相对更高(陈婷，孙晓敏，2016)。另外，共享信息偏差的出现受团队信息分布、决策任务特征、成员特征和信息分享动机等因素的调节。共享信息理论提示我们，在人机协同决策中，通过增加人机非共享信息的使用率，或可显著提升人机协同决策绩效。

(2)团队共享心智模型

共享心智模型(Shared mental model)指团队成员关于团队任务、成员关系、作业情境等的共同知识结构和认知框架。共享心智模型是团队协作的认知基础，在此基础上，团队成员彼此互动，从而完成任务目标(Cannon-Bowers et al., 1993)。共享心智模型由两部分构成，一是任务相关模型，包括关于设备、技术与任务的心智模型；二是团队相关模型，包括关于团队关系与团队交互过程的心智模型。共享心智模型理论认为，彼此一致的心智模型是增强团队成员适应性、团队交互有效性和团队绩效表现的重要保障。研究表明，团队心智模型的相似性程度可有效预测团队协作绩效，成员间心智模型的一致性越高，团队绩效表现相对越好(Lim & Klein, 2006; Mathieu et al., 2000)。Mohammed 和 Dumville (2001)进一步指出，除了认知结构相似性，团队成员在价值观、态度与信念等方面的相似性也应被视为共享心智模型的重要组成部分。

随着团队形式的扩展，共享心智模型亦被运用于“人类−智能体”团队中(Demir et al., 2020)。Jonker 等人(2010)将“人类−智能体”团队的共享心智模型定义为人机双方在心智模型上的相似程度。Fan 和 Yen (2010)发现，通过协助智能体更为准确地评估人类同伴的实时认知负荷，可显著改善人机心智模型的共享程度，进而提升人机团队的信息交流效率。另外，Scheutz 等人(2017)提出了“过程监控−共享心智模型−任务绩效”的综合分析框架，用以分析智能体(如机器人)的心智模型及其在提升“人类−智能体”混合团队绩效上的作用。共享心智模型理论提示我们，人类决策者与智能体在认知图式、态度、价值观等方面的相似性或是影响人机协同决策质量的重要因素。

(3)团队决策认知过程的 STC 机制

何贵兵(2002)认为，团队决策本质上是以团队交互为基础的团队认知适应性变化过程。决策开始后，团队成员首先按各方需求定向分享信息和知识(分布式分享，Distributive sharing)；随后，依据决策任务要求，不断将他人所分享的信息和知识纳入自身的心智模型，使之发生适应性转换和更新(适应性转换，Adaptive transformation)；之后，若团队的认知储备相较于决策任务要求仍有缺失，团队成员将开展更为深入的认知交互，构建形成现有团队知识库中未有的新概念与新联结(交互式构建，Interactive construction)。上述认知变化过程被称为团队认知的 STC (Sharing-Transformation-Construction)演化机制(何贵兵，2002)。以此模型为基础，管文颖(2006)、黄德斌(2007)通过 Tinsel Town 等团队投资决策任务，先后考察了领导风格、时间压力和反馈方式等对团队 STC 演化及最终决策绩效的影响。团队认知的 STC 演化理论提示我们，人机协同决策或需经历人机间知识的分布式分享、适应性转换与交互式构建等过程，以实现高认知兼容性。上述过程能否顺利展开取决于人机初始知识分布、人机交互方式、决策任务特征等因素。

**2.2　关于人机协同关系的研究**

(1)人机兼容性理论

人机兼容性(Human-agent compatibility)指人类用户与智能体在工作方式、交互方式、认知方式等方面的相互匹配与适应程度(Flemisch et al., 2008)，由外部兼容性(Outer compatibility)与内部兼容性(Inner compatibility)两部分构成(Bubb, 1993)。外部兼容性主要关注人类用户的感觉−运

动器官与智能体外部硬件接口的匹配程度(即人机外部交互界面的匹配程度); 内部兼容性主要考察人类用户与智能体在内部系统上的匹配程度。另外, Coll 和 Coll (1989)提出了人机“认知兼容性” (Cognitive compatibility)概念, 主要反映人类与智能体在信息处理方式上的匹配程度。Coll 和 Coll 认为, 假若智能体可依照人类的思维方式与目标用户展开互动, 用户对其可用性评价和接受意愿将显著提升。

研究表明, 在决策支持领域, 人机兼容性, 尤其是人机认知兼容性程度, 是影响人机决策绩效的重要因素(Burton et al., 2020)。然而, 由于人类决策者通常使用直觉式的信息加工方式, 而智能体主要采取分析式的信息加工方式, 因此, 只有人机双方共同做出相向调整, 人机认知兼容性与人机协同决策质量才能得到保障。具体而言, 智能体设计者在开发过程中需更多了解人类信息加工方式, 适度采纳基于启发式的决策原则, 并尽可能增加决策算法的透明度, 以提高人类用户对其建议的接受意愿(Hafenbrädl et al., 2016)。另一方面, 由于人类在决策过程中时常出现过度自信(Sieck & Arkes, 2005)或过度保守(Lim & O'Connor, 1996) 的情况, 因此, 人类决策者需正视自身的认知偏见并做出相应改变, 以提升人机认知兼容性, 保障人机协同决策的顺利展开。人机兼容性理论提示我们, 人机兼容性, 尤其是人机内部兼容性, 是影响人机协同决策绩效的关键因素。

(2)人机信任理论

在“人-人”工作团队中, 信任是人际可持续关系的基础, 建立信任可有效提升团队成员的工作态度与行为绩效(Dirks & Ferrin, 2001)。伴随智能技术的快速发展, 组织的智能化趋势日益明显, “人-机”工作团队不断涌现, 智能体的角色逐渐由“被动”工具向“主动”成员转变(Mittu et al., 2016)。诸多研究表明, 人类对智能体的信任程度显著影响其与智能体的交互, 如是否接受智能体提供的数据信息与决策建议、是否愿意与智能体共同完成工作任务等(Hancock et al., 2011; Lee & See, 2004)。因此, 理解人机信任的建立与维持过程, 对提升人机协同决策绩效具有重要意义。

目前, 有关人机信任的研究主要关注人类对智能体能力(competence)和可靠性(reliability)等方面的知觉。影响人机信任建立与维持的因素主要包含三类, 即人类因素、机器因素与环境因素(Hancock et al., 2011; Schaefer et al., 2016)。其中, 人类因素主要指与人类用户相关的一系列能力与特质, 如自我效能感、信任倾向等; 机器因素指与智能体相关的若干性能与属性, 如可靠性、错误率、透明度等; 环境因素指与人机互动相关的一系列团队与任务特征, 如任务重要性、任务负荷、人机责任分布和依存性等。三类因素共同决策人类用户对智能体的信任程度。此外, Hancock 等人(2011)的元分析结果表明, 机器因素对人际信任的影响相对较大, 环境因素的影响中等, 人类因素的影响则相对较小。人机信任理论提示我们, 人类决策者对智能体的信任是人机协同决策得以开展的前提, 信任建立受人类特征、智能体特征、环境特征等因素的影响。

(3)技术接受模型

Davis 等人(1989)所提出的技术接受模型(Technology acceptance model)以理性行为理论(Theory of reasoned action)为基础, 用于解释人类用户对信息技术的接受或拒绝倾向。该模型认为, 人类用户的行为意图影响其对特定技术的使用, 而行为意图取决于其对该技术的态度倾向。用户对技术的态度倾向受两方面因素影响: 其一为感知有用性(Perceived usefulness), 即用户是否认为该技术能显著提高当前任务绩效; 其二为感知易用性(Perceived ease of use), 即用户是否认为该技术易于使用。两者共同决定用户对特定技术的态度倾向与使用意图。假若某技术的感知易用性较低, 即使其感知有用性相对较高, 人类用户对其的使用意图亦较为有限。此外, 感知易用性可影响用户对感知有用性的评估, 且两者共同受到若干外在因素的作用, 如系统特征、用户特征、任务特征、组织特征、管理特征等。此后, Davis 和 Venkatesh (1996)以相关研究为依据, 对技术接受模型做了进一步修订。他们提出, 感知有用性与感知易用性可直接作用于个体的技术使用意图, 进而影响其技术使用行为。技术接受模型提示我们, 人类决策者对智能体有用性与易用性的感知将决定其对智能体的使用意愿与互动行为, 进而影响人机协同决策的最终绩效。

## 2.3 以往研究不足

以往关于“人-人”团队决策的研究主要关注团队共享信息、共享心智模型、团队决策认知过

程演化等内容，针对“人−机”协同关系的研究则主要聚焦于人机兼容性、人机信任、技术接受意愿等议题。相关研究表现出以下特征。

首先，有关人机协同决策的理论探索与实证支撑有待加强。现有的团队决策模型，如共享信息理论、共享心智模型、团队决策认知过程的 STC 机制等，主要基于对“人−人”团队决策的研究。随着组织智能化浪潮的兴起，组织的管理决策职能逐渐由人类成员与智能体共同承担，关于“人−机”协同决策形成机制与影响因素的研究尤为迫切并受研究者关注。虽然有学者尝试将“人−人”团队决策理论运用于“人−机”团队情境，并取得了初步研究成果，但考虑到“人−机”交互与“人−人”交互之间存在的重要差异，以上推广是否合理仍然存疑。因此，本项目尝试重新构建人机内部兼容性理论框架，并以信息利用协同、方案形成协同、方案评估协同与决策制定协同为主要维度，深入考察人机协同决策的形成过程与相关影响因素。

其次，现有人机兼容性理论的完备性仍有不足。人机兼容性理论认为，人机内部兼容性不足，尤其是人机认知兼容性缺乏，是引发人机协同决策问题(如信任度低、可控性低、透明度低、协同性低等)的关键原因；调整人类决策者与智能体的认知加工方式，使两者更具认知兼容性，是提高人机协同决策绩效的关键途径。然而，基于对“事实前提”和“价值前提”共同影响决策的认识以及对智能组织中人机协同决策过程的多角度分析，我们认为，除了人机认知兼容性，情感兼容性与价值兼容性亦是人机内部兼容性的重要组成部分，三类兼容性共同对人机协同决策的质量、效率及主观体验等方面产生实质性影响。因此，本项目拟拓展人机内部兼容性内涵，构建包含认知兼容性、情感兼容性与价值观兼容性的人机内部兼容性理论框架，深入探讨三者对人机协同决策的可能影响及其作用机制。

最后，基于理论逻辑和实证依据提出人机协同决策优化方法的研究有待加强。智能组织中的一系列重要决策正逐渐由人类成员与智能体共同完成。因此，如何提升人机协同决策绩效是一项极具现实意义的研究课题。从理论逻辑看，我们认为，人机兼容性尤其是人机内部兼容性，影响人机协同过程，并进而影响人机协同决策质量、决策效率和决策体验。然而这一影响机制仍需更多实证研究证据支持，特别是情感兼容性和价值兼容性的效应仍需检验。本项目拟通过系列现场研究和实验室模拟实验，揭示“内部兼容性−人机协同行为−人机协同决策绩效”三者关系，并考察人、机、团队、任务等因素的影响。在此基础上，从提升人机内部兼容性入手，提出若干改善人机协同决策绩效的方法，并在真实的组织场景中检验其有效性。这些探索，有助于开拓优化人机协同决策过程和决策绩效的新途径，也能为决策智能体设计者提供启示，从而有利于推进组织决策智能化进程。

## 3　研究构想

以往关于人机内部兼容性的研究主要探讨人机认知兼容性在人机协作任务中的作用。本项目认为，除了认知兼容性，情感兼容性与价值兼容性亦是人机内部兼容性的重要组成部分。其中，情感兼容性指人类与智能体对特定情感事件所做反应的匹配程度，以及人类对智能体作为协作伙伴的信任和认同；价值兼容性指人类与智能体在决策选择中遵循的价值观、道德原则等的匹配程度(Yokoi & Nakayachi, 2021)。以往关于情感兼容性的研究主要集中于人机信任，并认为人机信任有利于提高人类对智能体所提建议的接受度及协作意愿(Hancock et al., 2011)，但对人机情感反应一致性如何影响协作过程和协作绩效的研究仍相对缺乏。此外，关于价值兼容性是否及如何影响人机协同过程和协同绩效，还未见实证研究。我们假设，由上述三类兼容性所构成的内部兼容性是人机协同的重要基础，对人机协同决策的质量、效率和体验会产生实质性影响。因此，本项目拟拓展人机内部兼容性内涵，构建包含认知兼容性、情感兼容性与价值观兼容性的人机内部兼容性理论框架，深入探讨三者对人机协同决策的可能影响及其作用机制。

本项目拟结合决策心理学、认知科学、组织行为学等多学科理论与方法，通过现场访谈、问卷调查、实验室模拟与现场实验等手段，揭示智能组织人机协同决策中可能存在的问题及成因(研究一)；并从认知兼容性、情感兼容性、价值观兼容性三方面出发，考察人机内部兼容性对人机协同决策的影响方式与作用机制(研究二)；进而

提出若干可提升人机协同决策绩效的优化方法，并在智能组织决策场景中检验其有效性(研究三)。

### 3.1 研究一：人机协同决策主要问题及其成因的现场研究

研究一拟在现实的智能组织中考察人机协同决策可能存在的问题及原因。为此，首先需开发关于人机内部兼容性与人机协同行为的测评工具(子研究 1)，并将其运用于企业现场研究，从而了解智能体在组织场景中的使用现状、人类员工和决策者对智能体的使用意愿，以及人机协同决策中可能存在的问题与成因(子研究 2)。

对人机内部兼容性与人机决策协同性的有效测量是本项目研究的重要基础。然而，目前关于两者的测评工具仍相对缺乏。因此，子研究 1 尝试以认知兼容性、情感兼容性、价值兼容性为理论框架，开发《人机内部兼容性量表》；以信息收集协同、方案形成协同、方案评估协同与决策制定协同为主要维度，开发《人机决策协同性量表》。随后，子研究 2 将利用本项目开发的《人机内部兼容性量表》与《人机决策协同性量表》，结合现场访谈、问题调查、量表测试、个案分析等方法，探讨智能体在现实组织中的使用现状、人类员工对其的使用意愿，以及人机协同决策中存在的问题及其成因。

### 3.2 研究二：基于人机内部兼容性的人机协同决策机制研究

当前，关于人机内部兼容性如何影响人机协同决策过程和绩效的研究相对缺乏。研究二拟采用实验室模拟和现场研究等方法深入考察人机内部兼容性与人机协同决策间的内在关联。因此，研究二中的四项研究分别考察认知兼容性(子研究 3)、情感兼容性(子研究 4)和价值兼容性(子研究 5)对人机协同决策绩效(决策质量、决策效率、决策体验)的可能影响，并检验若干决策任务特征在其中的调节作用(子研究 6)。

子研究 3 主要关注认知兼容性如何影响人机协同决策绩效。人机认知兼容性指人类决策者与智能体在信息储备和信息加工等方面的一致性程度，主要受人机协同决策系统中双方的信息分布、加工方式、加工能力等因素的影响。因此，子研究 3 拟通过实验室模拟决策任务，考察信息量及其分布以及人机信息加工方式、加工能力对人机协同决策绩效的影响，并检验认知兼容性在其中的中介作用。子研究 4 主要考察情感兼容性在人机协同决策形成过程中的可能作用。人机情感兼容性指人类决策者对智能体的信任、认同和共情程度，其强弱主要取决于人类决策者对智能体能力、可靠性等因素的知觉。因此，子研究 4 将在实验室模拟决策任务下，探讨人类决策者对智能体内外特征的知觉如何形塑其与智能体的情感兼容性，并进而影响人机协同决策绩效。子研究 5 重点考察价值兼容性对人机协同决策的可能影响。人机价值兼容性指人类决策者与智能体算法在价值取向、道德准则等方面的一致性程度，其高低受人机价值取向、保护性价值观等因素的制约。因此，子研究 5 拟采用实验室模拟决策方法，考察人机价值取向、保护性价值观等因素对价值兼容性以及人机协同决策绩效的影响。子研究 6 尝试探讨决策任务特征等在人机兼容性影响人机协同决策中的调节作用。我们认为，决策任务的若干特征，如任务依存性、不确定性等因素或可改变人机兼容性对人机协同决策绩效的影响强弱。因此，子研究 6 拟在真实的组织决策场景下，检验任务依存性、决策不确定性等因素对人机兼容性与人机协同决策绩效间关系的调节作用。

### 3.3 研究三：基于人机内部兼容性的人机协同决策优化方法研究

基于研究一和研究二取得的成果，研究三拟通过现场研究，检验若干优化方法，如增加智能体决策透明度(子研究 7)、增设智能体决策结果反馈(子研究 8)等对提升人机内部兼容性和人机协同决策绩效的有效性。

智能体决策透明度指智能体所采用的决策规则向人类用户开放、为人类用户理解的程度。我们认为，当智能体的决策透明度得到保障时，人机之间的内部兼容性与协同决策质量亦将明显改善。另外，基于以往研究，我们预期，通过设置智能体决策结果反馈方式亦可增进人机内部兼容性，进而提升人机协同决策的整体绩效。子研究 7 与子研究 8 拟在现场实验中对上述假设的可靠性分别予以检验。

## 4 理论建构与应用前景

伴随智能技术的迅速发展，越来越多的企业试图将智能技术引入自身的各业务流程，智能化组织应运而生。在此背景下，人机协同决策逐渐

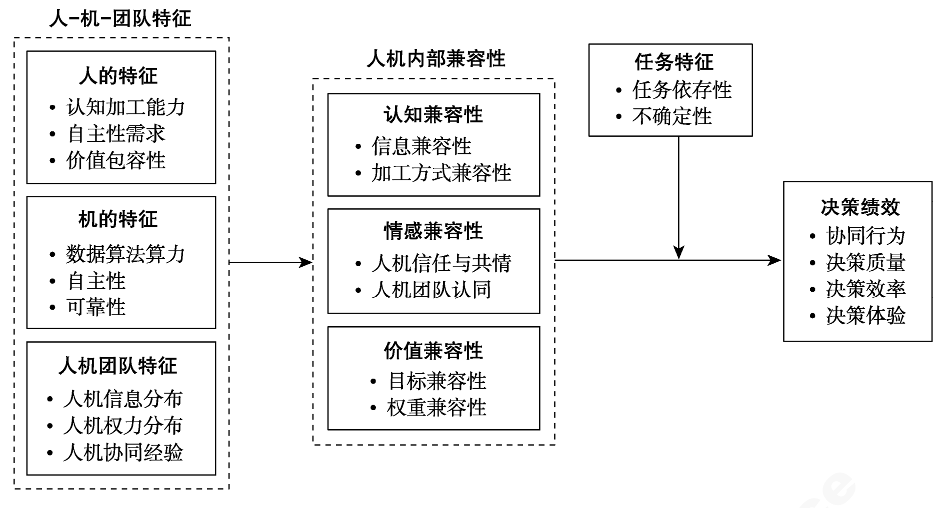



图 1　人机协同决策的内部兼容性理论框架

成为组织中新的决策方式。然而，当前的人机协同决策仍存在信任度低、可控性低、透明度低、协同度低等问题，其决策质量、效率和主观体验均有待提升。本项目拟以人机内部兼容性为分析视角，提出“人机认知兼容性、情感兼容性、价值兼容性是影响人机协同决策绩效的关键因素”这一崭新理论观点，在此基础上，采用现场访谈、问卷调查、个案研究、实验室模拟、现场实验等方法，揭示智能组织中人机协同决策存在的问题及成因，挖掘人机内部兼容性对人机协同决策的影响机制，并提出若干可有效改善人机协同决策绩效的优化方法(见图 1)。

本项目的研究具有理论创新性。以往关于人机内部兼容性的研究主要考察人机认知兼容性对人机协作行为的影响。我们认为，除认知兼容性外，情感兼容性与价值观兼容性亦是人机内部兼容性的重要组成部分，它们共同对人机协同决策绩效产生实质性影响。因此，本项目从三类内部兼容性理论框架出发，深入考察其在人机协同决策过程中的作用。该研究视角有望为未来人机协同决策研究提供新方向。

同时，本项目研究成果也具有良好的应用前景。本项目从增强人机认知兼容性、情感兼容性、价值兼容性角度切入，提出优化人机协同决策过程、提升人机协同决策绩效的方法，并在现场检验其有效性，有望为智能组织改进人机协同决策提供新途径，为决策智能体开发者提供新启示，进而有助于推进组织决策智能化进程。

## 参考文献


陈婷, 孙晓敏. (2016). 团队决策中的共享信息偏差:基于隐藏文档范式的机制、影响因素探究. *心理科学进展, 24*(1), 132−142.

管文颖. (2006). *团队决策中共享心理模型演化的影响因素研究* (硕士学位论文). 浙江大学, 杭州.

何贵兵. (2002). *群体动态决策的适应性行为及其内隐学习机制* (博士学位论文). 浙江大学, 杭州.

黄德斌. (2007). *团队决策中共享心理模型演化的 STC 机制研究* (硕士学位论文). 浙江大学, 杭州.

刘文川, 唐坚, 吴超. (2019). 人工智能与领导力: 基于复杂大数据支撑的政府决策机制研究. *改革与开放, 11*, 42−46.

许为, 葛列众. (2020). 智能时代的工程心理学. *心理科学进展, 28*(9), 1409−1425.

Bubb, H. (1993). Systemergonomie. In H. Schmidtke (Ed.), *Ergonomie* (pp. 333−420). München, Germany: CarlHanser.

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220−239.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Individual and group decision making: Current issues* (pp. 221–246). Lawrence Erlbaum Associates, Inc.

Chamorro-Premuzic, T., & Ahmetoglu, G. (2016). The pros and cons of robot managers. *Harvard Business Review, 12*, 2−5.

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in human performance and cognitive engineering*

*research* (Vol. 2, pp. 1−12). Bingley, UK: Emerald Group.

Coll, R., & Coll, J. H. (1989). Cognitive match interface design, a base concept for guiding the development of user friendly computer application packages. *Journal of Medical Systems, 13*(4), 227−235.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard business review*, *96*(1), 108-116.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982−1003.

Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. *International Journal of Human-Computer Studies, 45*(1), 19−45.

Demir, M., McNeese, N. J., & Cooke, N. J. (2020). Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies, 140*, Article 102436.

Dewhurst, M., & Willmott, P. (2014). Manager and machine: The new leadership equation. *McKinsey Quarterly, 4*(3), 76−86.

Dineen, B. R., Noe, R. A., & Wang, C. (2004). Perceived fairness of web-based applicant screening procedures: Weighing the rules of justice and the role of individual differences. *Human Resource Management, 43*(2-3), 127−145.

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science, 12*(4), 450−467.

Fan, X., & Yen, J. (2010). Modeling cognitive loads for evolving shared mental models in human-agent collaboration. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41*(2), 354−367.

Flemisch, F., Kelsch, J., Löper, C., Schieben, A., & Schindler, J. (2008). Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid, & K. A. Brookhuis (Eds.), *Human Factors for assistance and automation* (pp. 1−16). Maastricht, the Netherlands: Shaker Publishing.

Ghislieri, C., Molino, M., & Cortese, C. G. (2018). Work and Organizational Psychology Looks at the Fourth Industrial Revolution: How to Support Workers and Organizations? *Frontiers in Psychology, 9*, Article 2365.

Hafenbrädl, S., Waeger, D., Marewski, J. N., & Gigerenzer, G. (2016). Applied decision making with fast-and-frugal heuristics. *Journal of Applied Research in Memory and Cognition, 5*(2), 215−231.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors, 53*(5), 517−527.

Harms, P. D., & Han, G. (2019). Algorithmic leadership: The future is now. *Journal of Leadership Studies, 12*(4), 74−75.

Jonker, C., Van Riemsdijk, M., & Vermeulen, B. (2010, May). *Shared Mental Models: A Conceptual Analysis.* Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), Toronto, Canada.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15−25.

Lam, S. S., & Schaubroeck, J. (2011). Information sharing and group efficacy influences on communication and decision quality. *Asia Pacific Journal of Management, 28*(3), 509−528.

Langer, M., König, C. J., Gebhard, P., & André, E. (2016). Dear computer, teach me manners: Testing virtual employment interview training. *International Journal of Selection and Assessment, 24*(4), 312−323.

Larson, J. R., Christensen, C., Abbott, A. S., & Franz, T. M. (1996). Diagnosing groups: Charting the flow of information in medical decision-making teams. *Journal of Personality and Social Psychology, 71*(2), 315−330.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50−80.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society, 5*(1), 1−16.

Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior, 27*(4), 403−418.

Lim, J. S., & O'Connor, M. (1996). Judgmental forecasting with interactive forecasting support systems. *Decision Support Systems, 16*(4), 339−357.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science, 8*(4), 277−301.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273−283.

Mittu, R., Sofge, D., Wagner, A., & Lawless, W. F. (2016). *Robust intelligence and trust in autonomous systems.* Boston, MA: Springer.

Mohammed, S., & Dumville, B. C. (2001). Team mental

models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22*(2), 89−106.

Naim, I., Tanveer, M. I., Gildea, D., & Hoque, M. E. (2016). Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing, 9*(2), 191−204.

Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in human-machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior, 89*, 27−39.

Parry, K., Cohen, M., & Bhattacharya, S. (2016). Rise of the machines: A critical consideration of automated leadership decision making in organizations. *Group and Organization Management, 41*(5), 571−594.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 58*(3), 377−400.

Scheutz, M., DeLoach, S. A., & Adams, J. A. (2017). A framework for developing and using shared mental models in human-agent teams. *Journal of Cognitive Engineering and Decision Making, 11*(3), 203−224.

Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making, 18*(1), 29−53.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*(6), 1467−1478.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology, 53*(1), 81−93.

Wynne, K. T., & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science, 19*(3), 353−374.

Yokoi, R., & Nakayachi, K. (2021). The effect of value similarity on trust in the automation systems: A case of transportation and medical care. *International Journal of Human-Computer Interaction*, *37*(13), 1269−1282.

# Human-agent collaborative decision-making in intelligent organizations: A perspective of human-agent inner compatibility

HE Guibing, CHEN Cheng, HE Zetong, CUI Lidan, LU Jiaqi,
XUAN Hongzhou, LIN Lin

(*Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou* 310028, *China*)

**Abstract:** The era of artificial intelligence has already arrived. With the rapid development of intelligent technology, more and more companies are adopting this technology into their business processes to enhance their core competitiveness. Subsequently, human-agent collaborative work is becoming common, and human-agent collaborative decision-making (HACDM) is evolving as a new form of organizational decision-making. However, evidence shows that HACDM still faces challenges, such as low trust and controllability toward agents, low transparency of agents, and low collaboration between humans and agents. Therefore, how these challenges can be overcome to improve the decision quality, decision efficiency, and user experience of HACDM is crucial to the field of organizational decision-making. This project suggests that human-agent compatibility, especially human-agent inner compatibility (HAIC) which consists of cognitive, affective, and value compatibility, might be the fundamental factor affecting the performance of HACDM. Following the perspective of HAIC theory and using the multi-disciplinary methods from psychology, cognitive science, and organizational behavior, we intend to 1) reveal the existing problems within HACDM; 2) explore the impact of HAIC on the process and performance of HACDM; 3) propose methods to improve the performance of HACDM. This project's findings will contribute to the development of human-agent compatibility theory and human-agent collaboration theory, improve the performance of intelligent organizations and promote the intelligentization progress of HACDM.

**Keywords:** intelligent organization, human-agent collaborative decision-making, human-agent inner compatibility, cognitive compatibility, affective compatibility, value compatibility